Finnish National Bibliography Fennica as Linked Data

Osma Suominen
SWIB17, Hamburg
6 Dec 2017
THE NEEDS TO BE MORE WEBBISH.

BUT NOT TOO WEBBISH.

HOW LONG WILL THAT TAKE?

with apologies to Scott Adams
Why?

1. Making our data more visible, also internationally

2. Improving the quality and interoperability of our metadata

3. Building competency for the future

4. Why not? :)
1M bib records

125k person names

1M bib records

auth record

auth record

auth record

auth record

Image credit: MaryMaking blog
“From MARC silos to Linked Data silos”

“Family forest” of bibliographic data models, conversion tools, application profiles and data sets

As seen in:

SWIB16 talk
DCMI webinar
o-bib journal article
with separate Works and Instances like BIBFRAME, as enabled by the bibliographic extensions

**because it allows us to describe our resources from a common-sense, Web user perspective**

(and we get a metadata haircut for free!)

Special thanks to **Richard Wallis** for help with applying schema.org!
Publishing as Linked Open Data for human & machine access

Data dump downloads
- MARC records
- RDF N-Triples
- RDF HDT

Linked Data Fragments server
- LDF

RDF store
- Jena Fuseki

Flask app
- bib-lod-ui

Data formats:
- HTML+JSON-LD
- Linked Data RDF
- OpenSearch API

Linked Data RDF

SPARQL
Demo

http://data.nationallibrary.fi/bib/me/W00009584100

Spelunking UI...maybe?
## Our hairballs

### Index of /download

<table>
<thead>
<tr>
<th>Name</th>
<th>Last modified</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Directory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fennica-marc.zip</td>
<td>2017-12-05 00:23</td>
<td>254M</td>
<td></td>
</tr>
<tr>
<td>fennica.hdt</td>
<td>2017-12-05 00:23</td>
<td>492M</td>
<td></td>
</tr>
<tr>
<td>fennica.hdt.index</td>
<td>2017-12-05 00:23</td>
<td>293M</td>
<td></td>
</tr>
<tr>
<td>fennica.nt.gz</td>
<td>2017-12-05 00:22</td>
<td>739M</td>
<td></td>
</tr>
</tbody>
</table>

Apache/2.4.18 (Ubuntu) Server at data.nationallibrary.fi Port 80
NatLibFi Linked Data Fragments server

Available datasets

Browse the following datasets as Triple Pattern Fragments:

- **Fennica**: Fennica with an HDT back-end

The current dataset *index* contains metadata about these datasets.

Dataset index

Query dataset index by triple pattern

subject: __________________________________________

predicate: _______________________________________

object: _________________________________________

[Find matching triples]
Challenges
Work extraction

1. Extract works from MARC records
2. Create a work authority
3. Use and maintain it for cataloging
4. ???
5. Profit!

Not so easy in practice.
Lots of problems in the metadata that cause inconsistencies in the output.
Linking

- Person
- Subject
- Organization
- Work
- Place
- Instance

LCSH

Finnish Place Name Registry

Wikidata
“Cool URIs don’t change” -- Tim Berners-Lee

but we rely on conversion of MARC records that change all the timel
It’s open data.
Is it FAIR?
1. **Findable:**
   URIs as identifiers, with rich metadata

2. **Accessible:**
   URI lookup, SPARQL and LDF endpoints, downloadable data dumps

3. **Interoperable:**
   RDF representation using Schema.org and a little bit of RDAu

4. **Reusable:**
   CC0 license. Entities that are references also from other metadata
What next?

1. Enriching and cleaning the RDF data, e.g. using subclasses like Map
2. More links to other Linked Data sets
3. Expanding to new data sets: Viola discography, Arto article database
The Finnish Declaration of Independence was adopted by the Parliament of Finland on 6 December 1917
Thank you!
Questions?

osma.suominen@helsinki.fi - @OsmaSuominen

http://data.nationallibrary.fi - @NatLibFiData

Code:
https://github.com/NatLibFi/bib-rdf-pipeline
https://github.com/NatLibFi/bib-lod-ui

These slides: http://tinyurl.com/fennica-ld